下载APP

12 | 大促容量保障体系建设：怎样做好大促活动的容量保障工作（上）

2021-06-07 吴骏龙

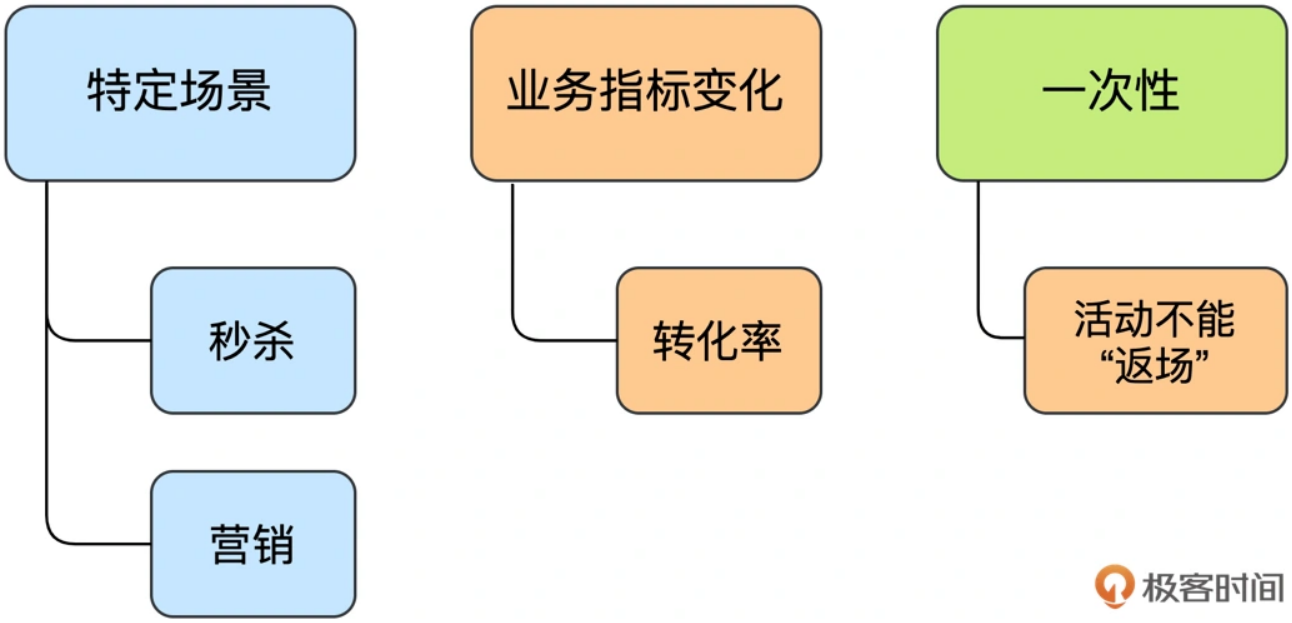
《容量保障核心技术与实战》 课程介绍 >

讲述：吴骏龙
时长 18:59 大小 17.40M

你好，我是吴骏龙。

在我的容量保障生涯中，有很多属于自己的记录：完成了上百次全链路压测、发现和改进了近千个容量问题、编写的压测平台总共支撑了近万亿次请求量……这些记录仍在延续。但在这些记录背后，也有我的一把“辛酸泪”：有未能发现容量隐患挨骂的，有把生产环境压挂了背锅的，有几次甚至都想撂担子不干了。但如果你问我什么时候压力最大，那我的答案只有一个，在大促期间做容量保障工作的压力最大。

为什么压力会那么大，因为大促活动和平时的业务场景很不一样，归纳一下，主要有三个特点。



首先，大促活动有一些日常没有的特定场景，比如秒杀活动、特殊的营销活动等，在没有常态数据可供参考的情况下，增加了容量保障的难度。现在一些大厂会倾向于复用之前的活动策略，比如支付宝的五福、红包雨等等，每年都是类似的套路。这除了有营销运营方面的考虑外，也是希望能有历史数据的沉淀，更好地预测用户流量。否则，每次都是新的玩法，容量保障的难度是非常大的。

其次，在大促场景下，一些业务指标会发生悄然的变化，比如转化率就是一个明显的例子，进入活动会场的用户数量较平时大幅增加，但这些新增的用户不一定都会真正下单，这样转化率就会变低，这时容量测试的目标也需要进行相应调整。

最后，活动往往是一次性的，无法返场，没有补救的机会。一旦出现问题，舆论方面可能都会有很大的压力。

这些特点都对容量保障工作提出了极高的要求。我们看一家公司的容量保障做得好不好，就看大促活动时它挂不挂，**大促活动是检验容量保障工作的一块试金石，非常考验团队的综合能力。**

我接触过不少容量保障工作的参与者，这其中包括性能测试人员、研发人员、运维人员、架构师等等，还有现在比较流行的 SRE（网站可靠性工程师），他们都在各自的岗位上为容量保障事业发光发热并小有成就，但容量保障是一张大图，尤其是大促容量保障更是多角色互动的协作过程，要达到组织一场大促容量保障的能力，需要具备很强的全局视角，这是这些角色暂时缺乏但渴望了解的。

当然，每家公司的业务模式不同，技术体系不同，技术栈不同，我在阿里本地生活牵头双 11 容量保障工作时，平均要对接 10 多个业务方、20 多项活动，横跨集团多个事业部，直接和间接参与到整个保障工作中的有上千人。我觉得单纯列举我经历的案例细节也许并不能直接复用到你所在的公司，可能你能看懂，但还是会有无从下手的感觉。

因此，在这一讲，我会将我在大促容量保障中的多年经验抽象成方法论，总结出一个通用的框架，告诉你大促容量保障应该要做哪些事情，帮你搭一个架子。这样，你可以结合所在公司的特点进行填空，甚至也可以对一部分框架进行取舍或延伸，**建设属于你的大促容量保障体系。**

明确背景与目标

背景和目标永远是前置项，我在接到一个大促容量保障任务的时候，首先不是去输出技术方案，而是找到产品和业务方，了解大促的业务背景，最好能拿到 PRD（产品需求文档）和具体的业务指标，先了解一下到底要举办哪些活动，活动的策略是什么，希望达到的效果是什么，活动周期和预估访问量是多少，等等。

为什么要知道这些信息呢？回忆一下我在“容量保障的目标”一节中谈到了的核心理念：**技术指标联系业务场景**。只有明确了业务场景，才能制定合理的技术目标。

然后，我才会去对接相关的业务技术团队，看一下每个活动的技术方案是怎样的，上下游调用链路有什么特点，并以此判断容量保障的侧重点，同时将业务指标转换为技术指标。

总之，先理解业务目标，再切入技术目标。

为了给你一些感性的认识，我以阿里本地生活的某年双 11 活动作为背景，分享一下在目标阶段我都做了些什么。建议你根据自己公司的特点，深入思考。

首先说下大背景，本地生活服务指的是在一个小范围的地区，比如一个城市，或者一个商圈范围内，将一些实体服务（餐饮、娱乐、医药、商超等）从线下搬到线上的服务过程。像饿了么订餐、口碑的到店扫码消费、叮咚买菜的送菜到家，都属于本地生活服务的范畴。这个背景一般公司内部成员都是清楚的。

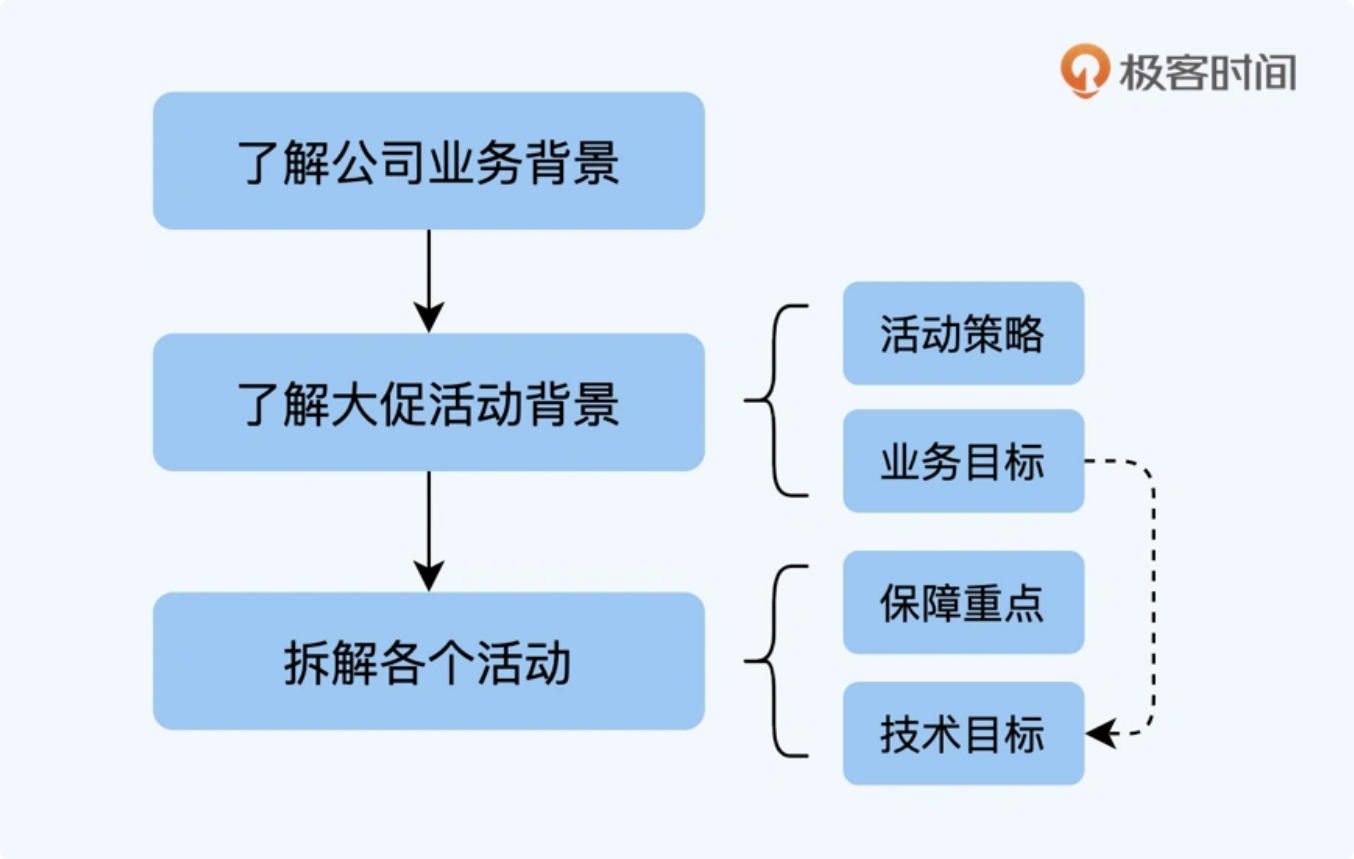
在高层确认双 11 活动战略后，我首先要去了解双 11 的业务目标。本地生活的双 11 大促活动业务目标主要是拉新和促进 GMV（网站成交总额），其中拉新的目标为新增 A 万用

户，GMV 目标为增长 B 万。同时，我还要了解业务场景，即双 11 会举办哪些活动，每个活动的细分业务指标是什么，比如双 11 当天会进行红包雨活动，预计将发放 10 万个面值 5 元的红包，等等。记住，**一定要获得量化的业务指标**，没有也要让业务方拍个脑袋出来，原因我后面会提到。

最后，针对各类活动场景，我会拆解出容量保障的侧重点，说白了，就是将业务目标转换为容量保障的技术目标，同样举几个例子：

1. 双 11 大促通过设置各种活动会场和页面，以吸引新用户参与，因此导购链路的流量会显著增高。这里的容量保障侧重点就是**导购链路的流量增长和转化率**。
2. 双 11 大促中，券和权益相关的营销场景会非常多，因为一个人一天不可能吃 10 顿饭，通过券和权益的形式，可以把这些购买力转移至后期消化，带动 GMV 增长。针对这个场景，应侧重考虑**券和权益涉及的新业务场景的流量预估**，这些场景在日常业务中是没有的，这时候就需要根据之前拿到的业务指标去换算，看到量化业务指标的重要性了吧。
3. 双 11 大促中，为制造一些轰动效应，会组织专场秒杀活动，或是类似“红包雨”这样的秒券活动。这里的侧重点就是**高并发场景的容量保障。**

到这里，我简单总结一下大促容量保障背景和目标的梳理思路，你可以看这张图。



明确了背景和目标后，我们具体展开大促容量保障的两项重要准备工作：**重点链路梳理和服务架构治理。**

重点链路梳理

大促活动一定会有一些重点链路，这些链路有的是新场景，有的是关键的老场景。归根结底，容量风险较大的链路，或是出现容量问题影响较大的链路都要涉及。我总结了以下几个侧重点：

1. 同步链路：同步链路指的是，链路上的服务是强依赖的，调用方需要等待被调用方执行完成后才能继续工作。在同步链路中，各服务的容量最容易互相影响，尤其是那些直接暴露给用户的入口接口，往往受制于众多下游接口的牵绊，需要协同看待。

比如，在双 11 活动中直接与买家交互的活动会场和导购页面，就是最靠近流量入口的同步链路。

2. 异步链路：异步链路与同步链路的概念正好相反，链路上的服务没有强依赖关系，调用方不需要等待被调用方执行完成，可以继续执行后续工作，被调用方在另一个线程内执行完成后，通过回调通知调用方。异步链路需要明确异步流量是从哪里过来的，异步流量的

量级有多大，在大促期间我方应用是否要做蓄洪，是否会由于消息重投而引起雪崩效应，等等。

比如，双 11 活动券的核销就是一个典型的异步链路，在大促期间，核销量可能会非常大，那么对于消息队列的容量保障就是重点。

3. 旁支业务链路：平时流量不大的“小”业务在大促场景下，流量是否会放大，是否会有叠加。有时候可能一个很不起眼的业务，在特定的场景下被反复调用后，会形成很高的终端延时。对于这类链路，主要是从业务场景的角度出发，细致地层层筛查，分析调用关系后得出结论。

比如，营销券底层服务，在平时流量不高，但双 11 期间大量活动和发券场景都会调用它，就会导致服务压力陡增。

4. 高并发链路：类似秒杀和红包雨这类链路，可能瞬间会产生极高的并发量，这类链路要尽量做到集群物理隔离，分层过滤流量，再通过容量测试去检验效果。

我需要提醒你注意，在这些要点中，高并发链路是尤其容易出问题的，一方面高并发链路往往会涉及新的活动场景，另一方面对高并发链路的保障工作涉及面较广，架构设计、容量治理和监控告警等任何一环出现问题都可能导致失败。**下面我以红包雨场景为你展开说明，如何来梳理这个高并发链路。**

红包雨活动是双 11 期间的一个典型的高并发场景，在每场红包雨期间，会有 10 秒的时长，在屏幕上落下数个红包，用户点击可以获取红包，红包雨下完后，将获取的红包结算给用户，用户可以在后续下单时使用这些红包。

在梳理这个高并发链路时，首先我们需要识别出**高并发的“爆点”** 在哪？也就是说，哪个环节，哪个接口，承受了最高的并发量。很显然，同一时间段会有大量用户点击领取红包，**计算用户是否中奖的接口**和**发放红包的接口**会承受较大压力，这就是爆点。

识别出爆点后，下面就要从两方面考虑：爆点的逻辑是否合理，能不能减少爆点的容量压力？以及如何设计容量测试方案。

针对爆点逻辑，需要思考是否每次都有必要计算是否中奖，可不可以事先计算好放到缓存中预热，减轻服务和数据库的压力；发放红包的接口是否可以异步发放，避免影响主流程，如果可行，在页面上需要加一些文案提示用户，比如：“您的红包将在 1 小时内发放到您的账户，请注意查收”。

确定了合理的逻辑后，就可以规划容量测试方案了，**方案需要尽可能与真实场景保持一致**，比如上面提到的，使用缓存预热的方式判断用户是否中奖，那么在容量测试时也需要提前按照相同的策略先把数据加入缓存，再进行测试。

最后，根据容量测试的结论，对红包雨整体链路进行合理的限流，预防容量事故的发生。

总结一下，在大促活动的链路梳理工作中，我给出了四类重要的链路，根据我的经验，做好这四类链路的容量保障，服务容量几乎可以确保万无一失。无论是哪一类链路，梳理工作的思路都是类似的，先识别其中的容量风险（爆点），再看能不能通过优化链路设计减少容量压力，并结合容量测试不断接近目标，最后通过限流等手段合理控制容量风险。

服务架构治理

除了链路梳理以外，对服务底层实现的关注也是很必要的，因为服务架构是影响服务容量的根本，在一个不合理的架构体系上进行容量保障是一件很痛苦的事情。从这个角度来说，架构的治理应该是要落到平时去做好的工作。只不过在大促活动中，受成本和时间约束，我们主要聊聊如何把这一工作进行得更高效些。

这里，我们将服务架构治理分为两部分，一部分是**新服务和新功能的架构治理**，另一部分是**针对存量服务的架构治理**。

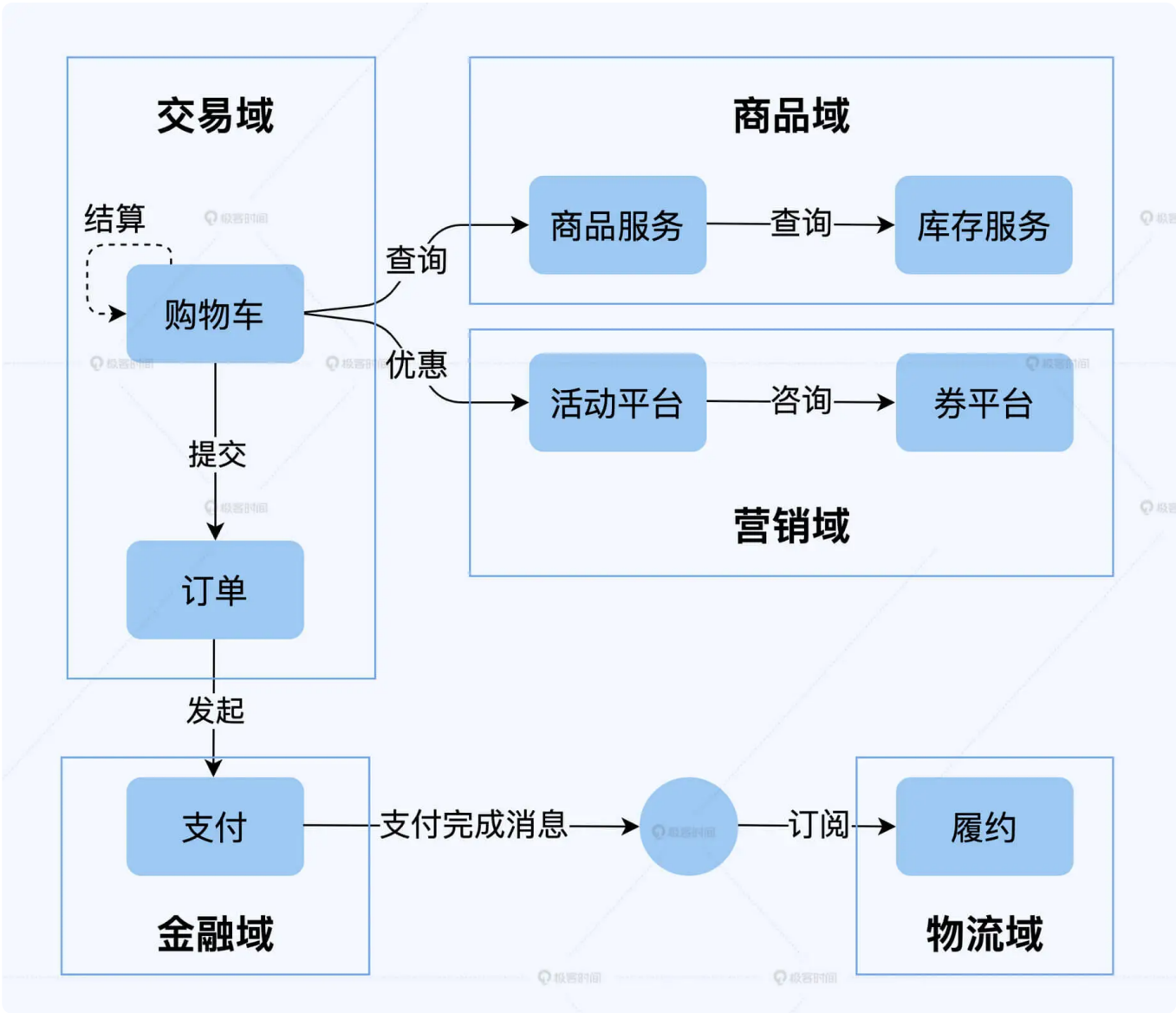
新服务和新功能的架构治理

对新服务和新功能的架构治理，倾向于通过绘制各种架构图（调用关系图、部署图、时序图，等等）的方式在架构评审时使用，以弱化文字。这样做的好处是，绘制这些图表能够倒逼你深入思考，特别是必须以体系化的思维方式去思考，能够在很大程度上避免低级的设计失误。此外，这些图表我们不规定必须以 UML（统一建模语言）的标准绘制，只要能表达清楚意义即可，重内容，轻形式。

下面我列举一些架构图表案例（为脱敏考虑，部分环节做了简化），以及为你解释一下它们的作用。

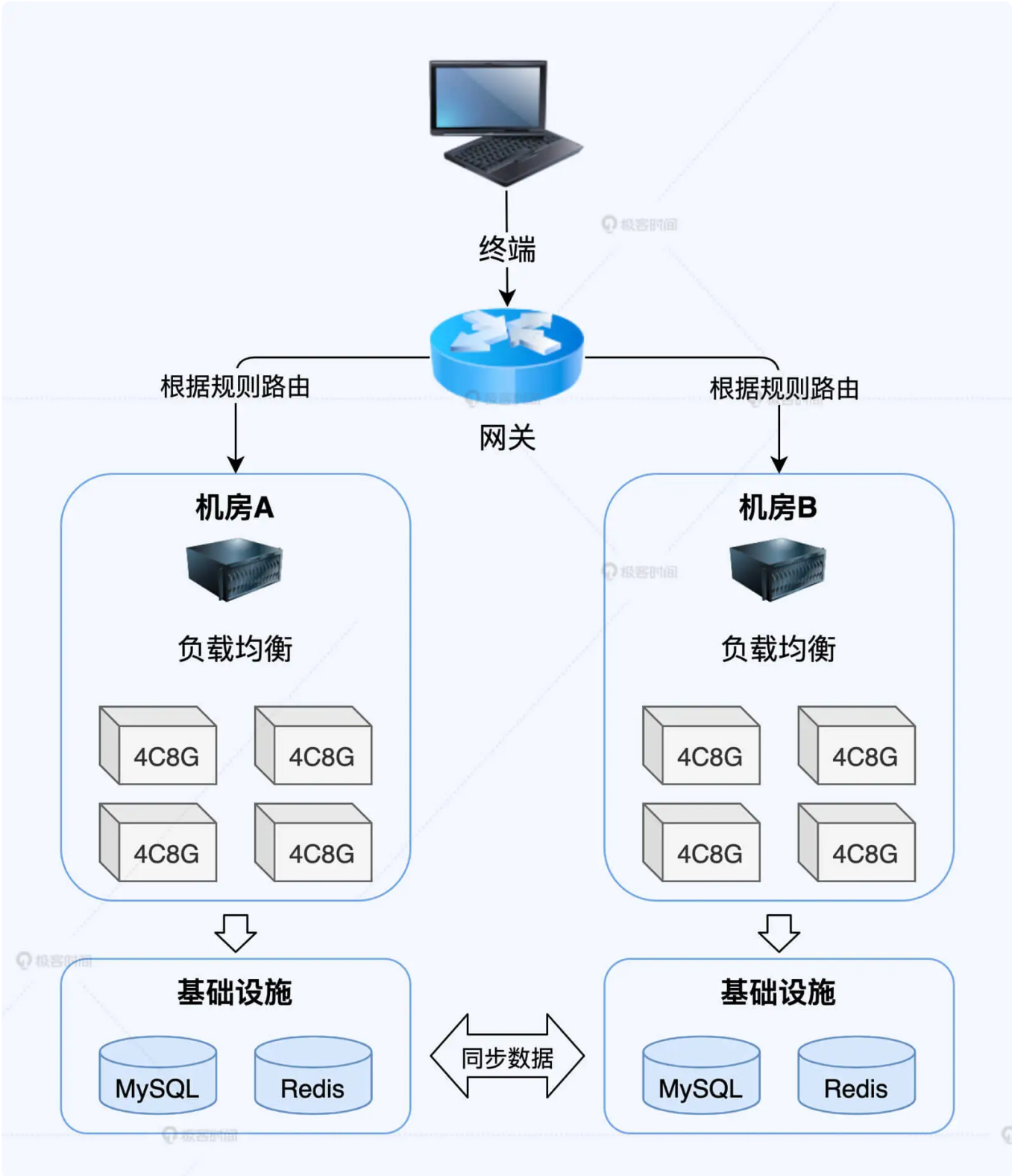
调用关系图：调用关系图表示服务之间主要模块的交互关系，以及与外部系统的交互关系。绘制调用关系图的目的是为了完整识别出服务的上下游和外部依赖，这是容量保障工作所需要的基础信息。

如下面的调用关系图所示，这是一个非常典型的电商业务形态的调用关系图，我对服务自身的调用进行了一定的简化，帮助你更清晰的看出跨领域的调用关系。整条链路经过了交易域、商品域、营销域、金融域和物流域五个领域，包含多个服务，获取到这些基础信息后，再进一步分析各个域之间的调用方式，和域内的细分链路，就不会有很大的信息遗漏。



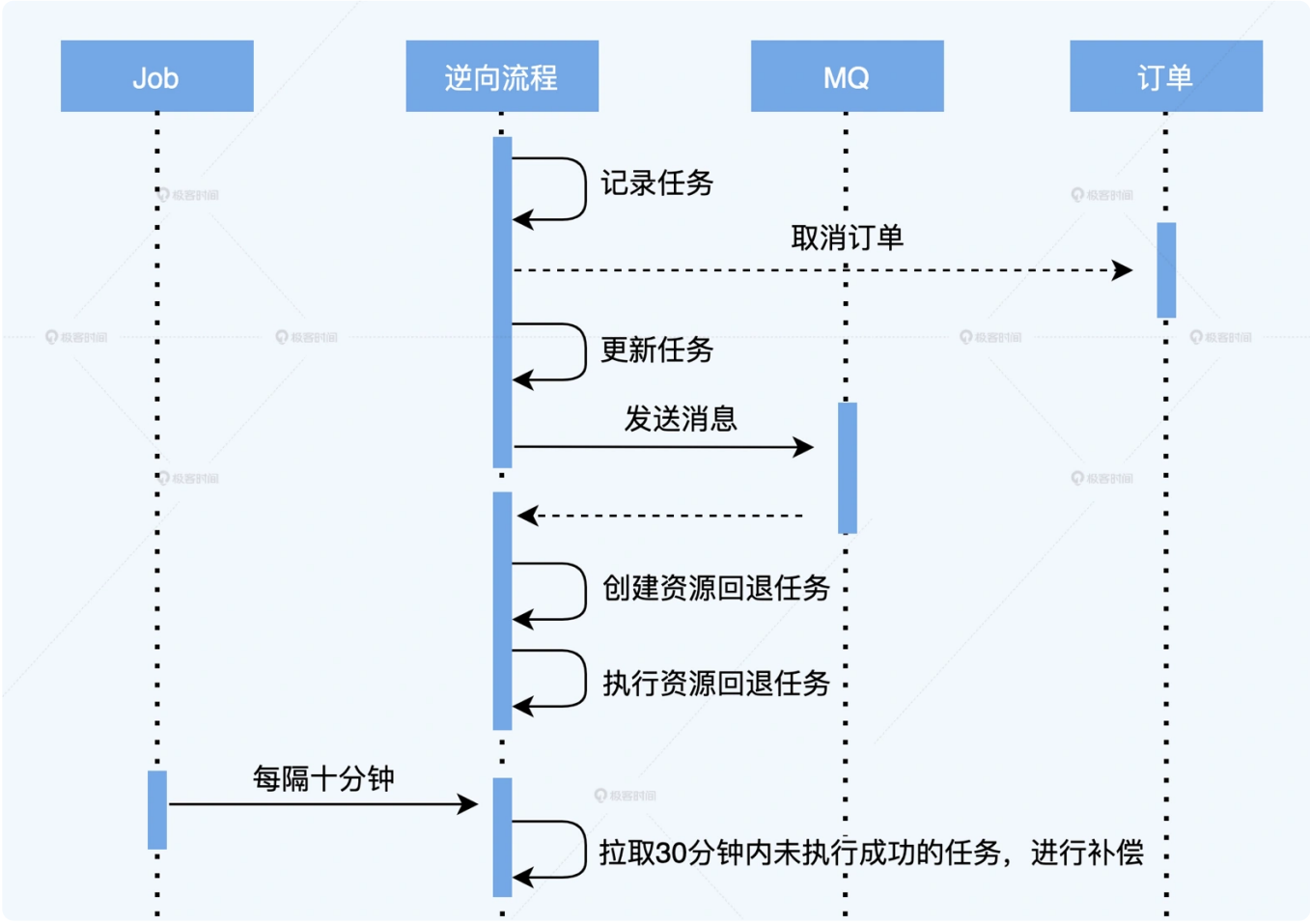
部署图：部署图描述服务部署的物理拓扑结构，在容量评估时我们需要知道服务的部署情况，来判断容量保障的范围。

比如下图为大促营销服务的部署图，可以看到服务是双活模式（双机房部署同时工作）部署的，服务之间存在一定的跨机房调用，这种情况下容量保障就需要考虑跨机房的专线带宽和延迟等因素。



时序图：时序图描述服务对象之间的调用顺序，它可以帮助我们快速识别出同步调用和异步调用，在链路梳理时很有用。

下图是交易逆向链路的简化时序图，可以很明显的看到，逆向流程中取消订单的操作，对订单服务的调用就是异步的，订单服务并没有立刻给出返回，而是由逆向流程自己创建一系列资源回退任务，异步执行回退，并由一个 Job 去做兜底补偿。在这个案例中，我们需要重点关注的就是 MQ 的容量。

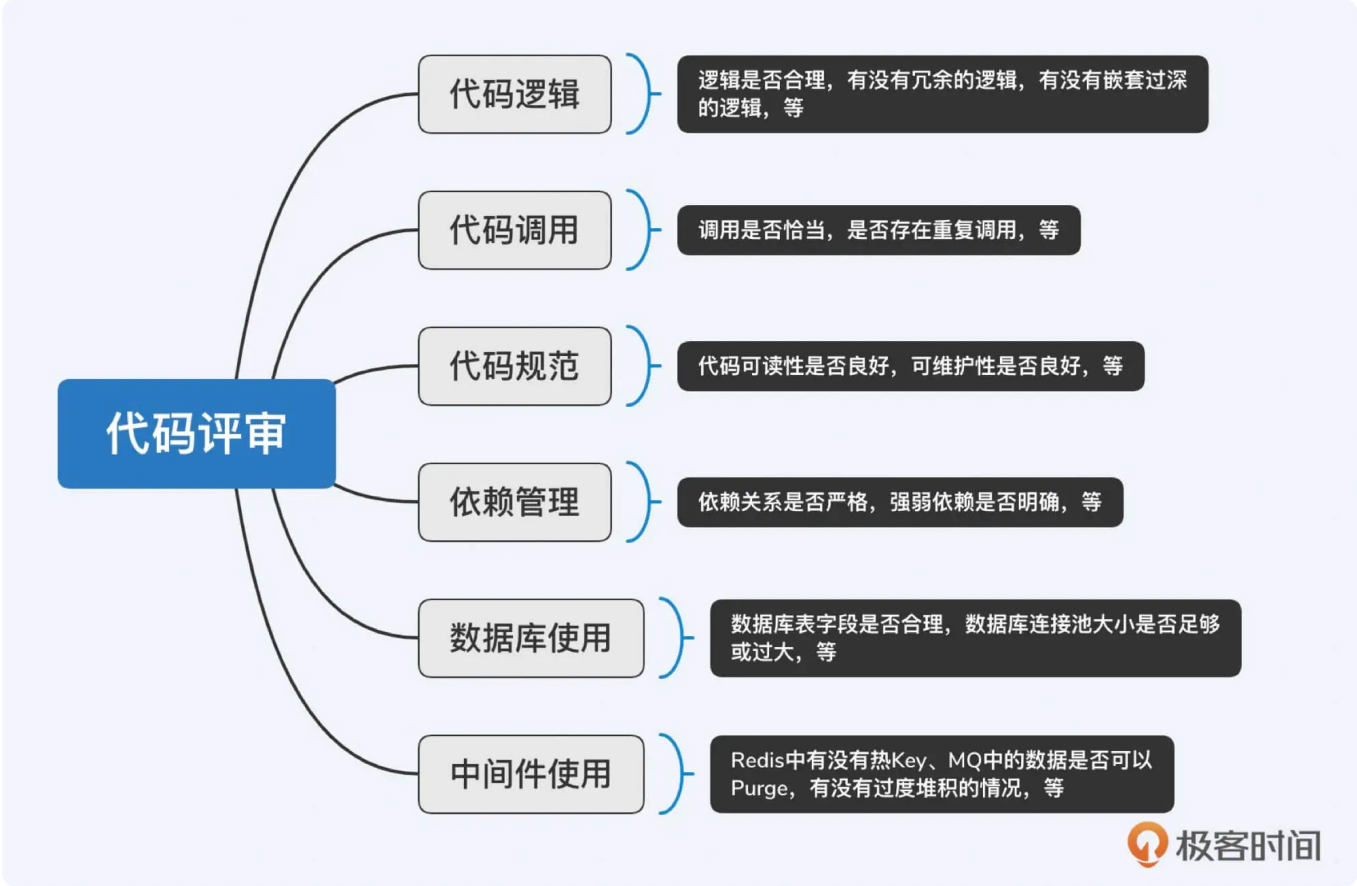


存量服务的架构治理

服务架构治理的另一部分是针对存量服务的治理，由于时间成本的约束，从性价比的角度考虑，可以优先梳理历史上发生过的事故和冒烟。这些历史风险如果在架构层面曾经存在隐患的，而且出事后没有得到有效解决的，就是定时炸弹，在大促前一定要从根上解决，防止层层打补丁，造成技术欠债越滚越多。

比如，某服务曾经发生过容量事故造成服务短时不可用的问题，原因是 Redis 使用时存在热 Key（大量请求访问某个特定 Key），那么我们就可以顺着这条线，排查一下其他服务是否也有类似的隐患，做到防患于未然。

另外，存量服务架构治理的另一个切入点是代码评审，通过对已有代码进行审查，发现潜在的隐患，同时制定改进策略。评审工作由架构师牵头，联合研发与测试一同进行，你可以按照以下脉络去实施。



下面这张表格是基于上述框架，我在大促期间推动服务架构治理中真实发现的具体问题和解决方案，供你参考。

问题	风险隐患	治理方案
调用方法存在snapshot包	修改后调用方无感知	更新为release包，强制检测
RedisLock分布式锁使用场景不明确	死锁、跨机房失效、大部分场景不触发	不滥用，明确场景，使用乐观锁或事务替代
表结构字段存在json	内容无法明确，可维护性差，无法建立索引	优化表结构
对外接口参数没有限制最大长度	长度过长导致接口超时	添加长度限制
Redis、MySQL、RPC重复调用次数过多	网络超时	尽量使用批量操作，平衡好传输次数和大小
存在用数字代表业务含义的code	可读性、可维护性、可扩展性差	枚举化

针对存量服务的架构治理工作，我建议可以按业务领域打散到每个业务团队，以专项的形式进行，甚至是组织头脑风暴来尽可能全面的分析问题。在双 11 活动中，我们进行了一些专项试点，获得了不错的效果，有些隐藏很深的问题，或长期处于灰色地带无人问津的隐患（如上述提到的用数字代表业务含义的 code），都得到了重视和解决。

总结一下，大促服务架构治理，针对新服务和新功能，要善于使用架构图去进行全局思考；针对存量服务，可以基于历史上发生过的事故和冒烟，结合代码评审的方式，查遗补漏。

总结

这一讲，我为你构建了一个大促容量保障的框架，帮助你认识大促容量保障要做哪些事，以及如何做好这些事。

目标为先，大促活动的业务场景有其特殊性，我们需要相应调整容量保障的目标，使其能够满足活动举办时的实际需求。

重点链路梳理和服务架构治理，是两项大促容量保障期间的重要准备工作，将有限的时间投入在最核心的部分，基于历史风险进行分析，是我给到你的经验之谈。

希望文中的案例能够帮助你，在大促活动中更好的落实容量保障工作。

课后讨论

在服务架构治理板块中，我给了一些案例，请你也思考一下，目前你负责或参与的业务服务中存在哪些架构隐患，可能会影响到服务容量的？

分享给需要的人，Ta订阅后你可得 16 元现金奖励

赞 1 提建议

© 版权归极客邦科技所有，未经许可不得传播售卖。 页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 11 | 与时俱进：云原生下的容量保障新趋势

下一篇 13 | 大促容量保障体系建设：怎样做好大促活动的容量保障工作（下）

精选留言

写留言

> 由作者筛选后的优质留言将会公开显示，欢迎踊跃留言。